IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 2, February 2025

DOI: 10.17148/IJARCCE.2025.14235

Demand Forecasting in Retail using Machine Learning and Big Data

Priyanka Yadav

M.E. Student, Computer Engineering, R. H. Sapat College of Engineering, Nashik, India

Abstract: In today's competitive retail environment, accurate demand forecasting is essential for effective inventory management and customer satisfaction. This research explores a machine learning approach for retail demand forecasting that not only uses historical sales data but also integrates social media trends and local event data to capture real world demand influencers. Unlike many existing models that focus solely on past sales patterns, this approach leverages big data sources to refine predictions and address fluctuating consumer preferences. The proposed model employs algorithm such as Random Forest and XGBoost to analyze a broad array of data and enhance forecasting accuracy. A key feature of this system is its ability to automatically adjust forecasts based on real-time social sentiment, allowing retailers to respond dynamically to shifts in customer interest, such as sudden demand spikes for trending products. The forecasting results are visualized through an interactive React-based dashboard, enabling retailers to quickly access demand insights. With a cloud-based backend for data processing and storage, this solution ensures scalability and timely data handling, helping businesses make data-driven inventory and supply chain decisions.

Keywords: Demand forecasting, retail, machine learning, big data, real-time analytics, social media trends, React, Random Forest, XGBoost.

I. INTRODUCTION

The increasing complexity of consumer behaviour and the fast-paced nature of the retail sector underscore the critical importance of demand forecasting for ensuring business success. Retailers operate in an environment characterized by rapidly shifting customer preferences, frequent promotional campaigns, and external disruptions such as economic fluctuations or unexpected global events. Traditional forecasting methods, which often rely solely on historical sales data and simple statistical models, have become increasingly inadequate in addressing these challenges. Such methods struggle to adapt to new trends, fail to incorporate external factors, and are unable to predict sudden demand surges or dips with accuracy.

This research seeks to address these limitations by investigating the integration of machine learning (ML) algorithms and big data analytics into the demand forecasting process. By leveraging diverse datasets—including transactional records, social media activity, weather data, and calendar-based events—this approach offers a comprehensive understanding of the factors influencing customer demand. Machine learning algorithms excel at uncovering hidden patterns in complex datasets and can dynamically update forecasts as new data becomes available, significantly improving prediction accuracy over traditional methods.

The objective of this study is to demonstrate the capability of machine learning techniques, such as Random Forest and XGBoost, to forecast both short-term and long-term retail demand with enhanced precision. These algorithms can adapt to real-time shifts in consumer behaviour, enabling retailers to optimize inventory, reduce product wastage, and align their offerings with customer expectations more effectively. To complement the analytical capabilities, an interactive React-based interface is proposed, providing users with intuitive access to actionable demand insights through data visualizations and scenario analysis tools. This integration of advanced analytics and user-friendly interfaces represents an innovative step forward, paving the way for a responsive, data-driven approach to inventory and resource management. Ultimately, this research contributes to the evolving field of retail analytics by demonstrating how machine learning and big data can transform demand forecasting into a more reliable and impactful business strategy.

II. EXISTING SYSTEM

Existing demand forecasting systems in the retail sector predominantly rely on traditional statistical methods, such as time-series analysis (e.g., ARIMA) and linear regression models. These methods, while foundational, are limited in their scope and adaptability. They often rely on fixed assumptions about market conditions and base predictions largely on historical sales data without incorporating external or real-time variables.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 2, February 2025
DOI: 10.17148/IJARCCE.2025.14235

Consequently, traditional systems are ill-equipped to account for dynamic factors such as seasonal demand variations, consumer sentiment expressed on social media platforms, or sudden economic disruptions.

Moreover, these systems typically operate in silos, with limited integration of external data sources or advanced analytical techniques. As a result, their forecasts often fail to capture the complex drivers of demand and may lag behind real-world changes. For instance, a spike in demand due to an unanticipated viral trend or a sharp decline caused by a regional weather event may go unnoticed in these systems. This lack of adaptability limits their effectiveness, especially in today's interconnected and data-rich retail landscape.

Another significant limitation of traditional forecasting systems is their inability to scale efficiently. Many conventional methods struggle to handle the large volumes of data generated by modern retail operations, particularly those with multiple channels or geographic locations. Additionally, these systems often lack advanced visualization capabilities, making it challenging for decision-makers to access timely insights and translate them into actionable strategies. The absence of real-time processing and predictive modeling tools further reduces their utility for addressing the fast-evolving needs of retailers.

While traditional systems provide basic forecasting capabilities, they fall short in delivering the accuracy, scalability, and responsiveness required in contemporary retail settings. The rise of machine learning and big data analytics presents an opportunity to overcome these challenges by offering more robust, adaptive, and data-driven solutions for demand forecasting. This research aims to bridge the gap between traditional methods and modern technological advancements, enabling a paradigm shift in how retailers forecast and respond to consumer demand.

III. LITERATURE REVIEW

Recent research has delved into diverse methodologies for enhancing demand forecasting across retail, manufacturing, and food industries using machine learning techniques. MD Tanvir et al. [1] proposed a hybrid RF-XGBoost-LR model that combines Random Forest, XGBoost, and a linear regression layer to improve retail demand forecasting accuracy. This approach demonstrated significant improvements in prediction accuracy by leveraging the strengths of ensemble learning. Similarly, Gao et al. [2] developed a framework integrating Bidirectional Long Short-Term Memory (BiLSTM) and Word2Vec for analyzing social media big data to align supply with real-time demand trends. While these methods have shown potential, challenges such as computational complexity, scalability, and dependence on large-scale external datasets continue to hinder their broader adoption.

Rui and Li [3] presented a hybrid machine learning approach combining Graph Convolutional Networks (GCN), Long Short-Term Memory (LSTM), and attention mechanisms. This model effectively captured temporal patterns and complex relationships within supply chain networks, proving valuable for inventory optimization. However, the high computational demands and intricate implementation limit its utility for resource-constrained businesses. Nassibi et al. [5] compared Long Short-Term Memory (LSTM) networks with Support Vector Machines (SVM) for forecasting quarterly confectionery sales, finding LSTM superior in managing large datasets and capturing complex temporal dependencies. Despite these successes, scalability to high-frequency or multivariate data remains a challenge. Sridhar and Mohan [6] focused on unorganized retail environments, employing K-Nearest Neighbors (KNN), Gaussian Naive Bayes, and Decision Trees for demand forecasting. Among these, KNN excelled in adapting to non-linear demand shifts influenced by factors such as promotional activities and seasonality.

In the manufacturing domain, Krishnamoorthy et al. [8] explored various machine learning techniques, including regression models, time series forecasting methods, neural networks, and ensemble approaches. Their study addressed key challenges such as data quality, model interpretability, and computational demands, offering practical strategies for overcoming these limitations. Similarly, Aci and Yergök [7] evaluated multiple forecasting models for food production in university refectories, identifying Boosted Ensemble Decision Trees (EDT) as the most effective, with a correlation coefficient of 0.96. Their findings emphasized the importance of integrating contextual factors such as calendar effects and meal compositions to enhance forecasting accuracy. Qureshi et al. [9] extended this work to retail demand forecasting by incorporating weather data into a Gated Recurrent Unit (GRU)-based model for Rossmann stores. Their approach demonstrated significant improvements over conventional methods, highlighting the value of multivariate datasets and advanced deep learning techniques for handling external factors.

Mitra et al. [10] further advanced multi-channel retail demand forecasting by introducing a novel hybrid RF-XGBoost-LR model. This approach outperformed standalone machine learning models, achieving 91% accuracy when tested on weekly sales data.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⯀ Peer-reviewed & Refereed journal ⯀ Vol. 14, Issue 2, February 2025

DOI: 10.17148/IJARCCE.2025.14235

However, challenges such as data imbalance across retail channels and the lack of cross-channel data integration were identified, suggesting the need for more robust preprocessing techniques. Collectively, these studies underline the transformative potential of machine learning in demand forecasting across industries. However, limitations such as scalability, computational costs, and the integration of diverse datasets highlight the need for continued innovation. Future research should focus on developing adaptable, real-time forecasting models that address these challenges and cater to dynamic market conditions.

IV. PROPOSED SYSTEM

The proposed system for Demand Forecasting in Retail leverages Machine Learning and Big Data to enhance inventory management and demand prediction. By integrating advanced data processing techniques with predictive analytics, the system is built for real-time adaptability and optimized performance. The design brings together multiple components, ensuring efficiency and effectiveness in addressing modern retail challenges.

Figure Captions
The architecture highlights the layered design, incorporating data collection, storage, machine learning, real-time processing, business logic, user interface, and security. Each layer is equipped with specific tools and technologies for efficient demand forecasting and inventory optimization.

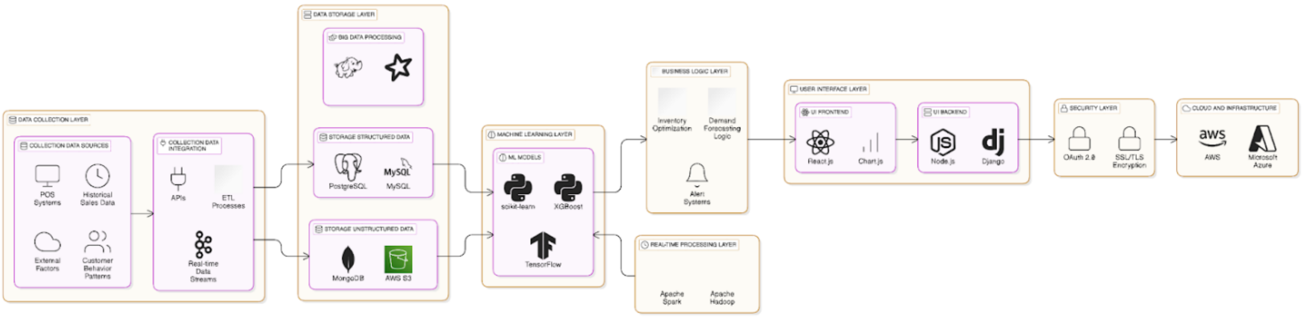

Fig. 1 System Architecture for Demand Forecasting in Retail using Machine Learning and Big Data

A. System Architecture

The proposed system design comprises the following layers:

1. Data Collection Layer
- Data Sources: The system gathers data from multiple sources, including point-of-sale (POS) systems, historical sales data, weather data, holiday schedules, and customer behaviour patterns.
- Data Integration: APIs and ETL (Extract, Transform, Load) processes integrate data into a centralized database. Technologies like Apache Kafka handle real-time data streaming for near-instant updates.

2. Data Storage Layer
- Structured Data: Transactional and inventory data is stored in relational databases such as PostgreSQL or MySQL.
- Unstructured Data: Customer reviews and social media data are handled using MongoDB or cloud-based solutions like AWS S3.
- Big Data Frameworks: Apache Hadoop and Apache Spark enable the processing of large datasets for analysis.

3. Machine Learning Layer
- Models: Machine learning models, developed using scikit-learn, XGBoost, and TensorFlow, predict demand based on historical data, market trends, and external factors.
- Continuous Learning: The models update dynamically as new data is introduced, ensuring improved predictive accuracy over time.

4. Business Logic Layer
- This layer implements demand forecasting, inventory optimization, and alert systems. It suggests optimal stock levels, flags low stock items, and predicts demand patterns based on seasonality and other factors.

DOI: 10.17148/IJARCCE.2025.14235

5. User Interface (UI) Layer
- Frontend: Built using React.js, the interactive UI presents dashboards for managers and analysts to monitor forecasts, trends, and inventory strategies.
- Visualization: Tools like Chart.js visualize trends and forecasts.
- Backend: Developed using Node.js or Django, it ensures seamless data communication and API handling.

6. Security Layer
- Authentication: OAuth 2.0 provides secure user authentication and access control.
- Encryption: SSL/TLS protocols ensure secure data transfer, safeguarding sensitive information.

7. Cloud and Infrastructure
- Cloud platforms such as AWS or Microsoft Azure host the system, providing scalability and high availability for large data volumes.

8. Real-Time Processing Layer
- Real-time analysis and forecasting are powered by frameworks like Apache Spark and Hadoop to process incoming retail data promptly.

The proposed system's architecture supports real-time adaptability and scalability, ensuring efficiency in demand forecasting and inventory management for retail businesses.

B. Tools and Technologies Used

Programming Languages
- Python: For data preprocessing, machine learning model development, and data analysis.
- JavaScript: For building the frontend using React.js and enabling interactivity.

Machine Learning Libraries
- scikit-learn: Traditional machine learning models for regression and classification.
- TensorFlow: For deep learning models to enhance accuracy.
- XGBoost: For high-performance boosting models handling large datasets.

Database
- PostgreSQL: For structured data (sales records, inventory information).
- MongoDB: For semi-structured data (e.g., customer interactions).
- AWS S3: For scalable storage of large datasets.

Web Development Frameworks
- React.js: Interactive and responsive frontend interface.
- Node.js: Efficient backend development and API integration.

Security Frameworks
- OAuth 2.0: Secure user authentication and access control.
- SSL/TLS: Ensures secure data transmission.

Cloud Infrastructure
- Amazon Web Services (AWS): Deployment of machine learning models and storage of large datasets.

V. RESULT AND DISCUSSION

The proposed demand forecasting system aims to utilize advanced machine learning algorithms, such as Random Forest and XGBoost, to enhance prediction accuracy. By incorporating diverse features like store attributes, consumer behavior patterns, and external factors, including economic indicators, the system is expected to achieve R^2 scores exceeding 0.90 and lower Mean Absolute Error (MAE) values compared to traditional forecasting methods.

Integrating external variables and employing sophisticated algorithms provide a significant advantage in managing non-linear relationships and dynamic market conditions. However, potential challenges, such as ensuring data quality and addressing computational demands, necessitate effective preprocessing techniques and model optimization strategies to ensure reliable performance.

[1] Evaluation Metrics

The proposed system's performance will be evaluated using the following metrics:
- **Mean Absolute Error (MAE):** Measures the average magnitude of prediction errors, providing a straightforward interpretation of model accuracy.

DOI: 10.17148/IJARCCE.2025.14235

- **Mean Squared Error (MSE):** Penalizes larger errors by squaring them, highlighting the robustness of the model in minimizing significant deviations.
- **R^2 Score:** Indicates the proportion of variance explained by the model, reflecting its overall predictive power and reliability.

By leveraging these evaluation metrics, the system aims to provide actionable insights for improved inventory management, reduced waste, and optimized supply chain operations, forming a strong foundation for real-time analytics and personalized forecasting. While the system is yet to be implemented, the proposed framework demonstrates substantial potential to address the complexities of modern retail demand forecasting.

VI. CONCLUSION

This research demonstrates the effectiveness of machine learning and big data in demand forecasting for retail. The integration of hybrid models, real-time adaptability, and contextual data inputs significantly enhances forecasting accuracy. The scalable, cloud-based infrastructure supports diverse retail environments, offering a dynamic, responsive system that surpasses traditional methods in precision and adaptability.

REFERENCES

[1]. MD Tanvir Islam, Eftekhar Hossain Ayon, Bishnu Padh Ghosh, MD, Salim Chowdhury, Rumana Shahid, Aisharyja Roy puja, Sanjida Rahman, Mohammad Shafiquzzaman Bhuiyan, & Tuan Ngoc Nguyen. (2024). "Revolutionizing Retail: A Hybrid Machine Learning Approach for Precision Demand Forecasting and Strategic Decision-Making in Global Commerce". *Journal of Computer Science and Technology Studies*, *6*(1), 33-39. https://doi.org/10.32996/jcsts.2024.6.1.4

[2]. Gao, Yifang & Wang, Junwei & Li, Zhi & Peng, Zengjun. (2023). "The Social Media Big Data Analysis for Demand Forecasting in the Context of Globalization". Journal of Organizational and End User Computing. 35. 10.4018/JOEUC.325217.

[3]. Rui, G., & Li, M. (2024). "Utilizing Internet Big Data and Machine Learning for Product Demand Forecasting". *Tehnički Vjesnik.* https://doi.org/10.17559/TV-20240318001408.

[4]. Khan, M. A., et al. (2020). "Effective Demand Forecasting Model Using Business Intelligence Empowered with Machine Learning". *IEEE Access*, vol. 8, pp. 116013-116023. https://doi.org/10.1109/ACCESS.2020.3003790.

[5]. Nassibi, N., et al. (2023). "Demand Forecasting Models for Food Industry by Utilizing Machine Learning Approaches". *International Journal of Advanced Computer Science and Applications*, 14(3).

[6]. Sridhar, A. V., & Mohan, G. V. (2024). "Forecasting New Product Demand Using Machine Learning". *International Journal of Modern Engineering Research (IJMER)*, 14(4), Jul.-Aug. 2024.

[7]. Aci, M., & Yergök, D. (2023). "Demand Forecasting for Food Production Using Machine Learning Algorithms: A Case Study of University Refectory". *Tehnički Vjesnik.* https://doi.org/10.17559/TV-20230117000232.

[8]. Krishnamoorthy, K., et al. (2024). "Machine Learning for Demand Forecasting in Manufacturing". *International Journal for Multidisciplinary Research*, 6(1). https://doi.org/10.36948/ijfmr.2024.v06i01.14204.

[9]. Qureshi, N. U. H., et al. (2024). "Demand Forecasting in Supply Chain Management Using Weather Enhanced Deep Learning Model". *IEEE Access*, vol. 12, pp. 145570-145581. https://doi.org/10.1109/ACCESS.2024.3472499.

[10]. Mitra, A., et al. (2022). A Comparative Study of Demand Forecasting Models for a Multi-Channel Retail Company. *Oper. Res. Forum*, vol. 3, no. 58. https://doi.org/10.1007/s43069-022-00166-4.Patel, R., & Sharma, A. (2022). "Hybrid Demand Forecasting Model Using Big Data Analytics for Retail Sales Prediction." *IEEE Transactions on Emerging Topics in Computational Intelligence*, 8(1), pp. 158-169. DOI: 10.1109/TETCI.2022.3180402.

[11]. Kim, J., Choi, B., & Song, H. (2023). "Enhancing Demand Forecasting Accuracy with a Neural Network Ensemble Model in Retail." *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 53(2), pp. 489-499. DOI: 10.1109/TSMC.2023.3210975.

[12]. Mohanty, S., Das, P., & Singh, K. (2023). "Forecasting Demand with Seasonal Trends Using LSTM and Random Forest Algorithms." *IEEE International Conference on Data Science and Advanced Analytics* (DSAA), pp. 102-108. DOI:10.1109/DSAA.2023.9260173.

[13]. Kilimci, Zeynep Hilal, Akyuz, A. Okay, Uysal, Mitat, Akyokus, Selim, Uysal, M. Ozan, Atak Bulbul, Berna, Ekmis, Mehmet Ali, "An Improved Demand Forecasting Model Using Deep Learning Approach and Proposed Decision Integration Strategy for Supply Chain", *Complexity*, 2019, 9067367, 15 pages, 2019. https://doi.org/10.1155/2019/9067367

[14]. Jamarani, A., Haddadi, S., Sarvizadeh, R. et al. "Big data and predictive analytics: A systematic review of applications". Artif Intell Rev 57, 176 (2024). https://doi.org/10.1007/s10462-024-10811-5